INL/CON-06-11953
PREPRINT

Bayesian Modeling of Time Trends in Component Reliability Data Via Markov Chain Monte Carlo Simulation

2007 ESREL Safety and Reliability Conference

D. L. Kelly

June 2007

The INL is a
U.S. Department of Energy
National Laboratory
operated by
Battelle Energy Alliance

Idaho National Laboratory

This is a preprint of a paper intended for publication in a journal or proceedings. Since changes may be made before publication, this preprint should not be cited or reproduced without permission of the author. This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, or any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use, or the results of such use, of any information, apparatus, product or process disclosed in this report, or represents that its use by such third party would not infringe privately owned rights. The views expressed in this paper are not necessarily those of the United States Government or the sponsoring agency.

Bayesian Modeling of Time Trends in Component Reliability Data via Markov Chain Monte Carlo Simulation

D. L. Kelly
Idaho National Laboratory, P. O. Box 1625, Idaho Falls, ID, USA 83415-3850, Dana.Kelly@inl.gov

ABSTRACT: Markov chain Monte Carlo (MCMC) techniques represent an extremely flexible and powerful approach to Bayesian modeling. This work illustrates the application of such techniques to time-dependent reliability of components with repair. The WinBUGS package is used to illustrate, via examples, how Bayesian techniques can be used for parametric statistical modeling of time-dependent component reliability. Additionally, the crucial, but often overlooked subject of model validation is discussed, and summary statistics for judging the model's ability to replicate the observed data are developed, based on the posterior predictive distribution for the parameters of interest.

1 INTRODUCTION

Much past work has been devoted to modeling failures with repair under the assumption that the repairs restore the component to "as good as new" condition, that is, under the assumption that the stochastic point process being observed is a renewal process. Disproportionately less work has addressed the perhaps more realistic assumption that repairs make the component "as good as old." And most of this work has been devoted to qualitative analysis and frequentist estimation. Under the assumption of a renewal process, the times between failures (interarrival times) are independently and identically distributed (iid), and this makes the statistical analysis straightforward. However, under the "as good as old" assumption for repair, the interarrival times are not iid; the distribution for the i^{th} time is dependent upon t_{i-1}. This paper focuses on Bayesian analysis of the "as good as old" assumption.

2 NONHOMOGENEOUS POISSON PROCESS (NHPP) FOR FAILURES

Under a homogeneous Poisson process (HPP), the number of failures, X, in time t is described by a Poisson distribution:

$$f(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!} \tag{1}$$

In Equation 1, λ is called the intensity of the process. For the HPP, λ is independent of time, and the expected number of failures in time t is given by λt. It is a well known result that the interarrival times of the HPP are iid exponential:

$$f(t) = \lambda e^{-\lambda t} \tag{2}$$

Relaxing the assumption of constant λ leads to the NHPP. The number of failures in time t is still Poisson-distributed, but the expected number of failures in time t is given by

$$E[X(t)] = \int_0^t \lambda(s)ds \tag{3}$$

Moreover, the expected number of failures in any given time interval, $[t_1, t_2]$, is given by

$$\int_{t_1}^{t_2} \lambda(t)dt \tag{4}$$

If λ(t) is increasing with time, the times between failures are decreasing with time; the component is aging. Conversely, if λ(t) is decreasing with time, the times between failures are increasing with time, and the component is experiencing reliability growth.

The functional form of $\lambda(t)$ must be specified in order for parametric analysis to proceed. Common forms for $\lambda(t)$ include the power-law process,

$$\lambda(t) = \frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1} \quad (5)$$

the loglinear model,

$$\lambda(t) = exp(a + bt) \quad (6)$$

and the linear model,

$$\lambda(t) = a + bt \quad (7)$$

This paper will focus on the power-law process, as it is mathematically convenient, has been the subject of some past analysis, and subsumes the constant model ($\alpha = 1$) and the linear model ($\alpha = 2$). The time to first failure for the power-law process has a Weibull distribution with shape parameter α and scale parameter β:

$$f(t_1) = \frac{\alpha}{\beta}\left(\frac{t_1}{\beta}\right)^{\alpha-1} exp\left[-\left({}^{t_1}\!/_{\beta}\right)^{\alpha}\right] \quad (8)$$

For this reason, the power-law process is sometimes referred to as a Weibull process. This name is unfortunate in that analysts have sometimes had the mistaken notion that a sample of interarrival times (or in some cases the times themselves!) from a power-law process is an iid sample from a Weibull(α, β) distribution. As pointed out above, this assumption is only valid if one is observing a renewal process.

3 BAYESIAN ANALYSIS OF POWER-LAW PROCESS

Relatively little work has been done on Bayesian analysis of a power-law process. Notable references are (Guida et al., 1989) and (Chen, 2004). A reason for the dearth of work in this area may be the relative intractability of the Bayesian approach. As noted by (Guida et al., 1989), "[Bayesian procedures] are computationally much more onerous than the corresponding maximum likelihood ones, since they in general require a numerical integration." This problem has been obviated by the advent of Markov chain Monte Carlo (MCMC) techniques and software for implementing such approaches.

3.1 *Likelihood Function*

This paper analyzes the case in which the observation process is failure-truncated. The alternative, in which the observations stop after a fixed time, is straightforward to analyze in a similar manner. As pointed out above, the time to first failure has a Weibull(α, β) distribution, given in Equ. 8. For i = 2, …, n, we must use the condition that the failure times are ordered:

$$f(t_i \mid t_{i-1}) = f(t_i \mid T_i > t_{i-1}) = \frac{f(t_i)}{Pr(T_i > t_{i-1})} \quad (9)$$

This is a truncated Weibull distribution. Thus, for i = 2, …, n, we have

$$f(t_i \mid t_{i-1}) = \frac{\alpha}{\beta^{\alpha}}(t_i)^{\alpha-1} exp\left[-\left(\frac{t_i}{\beta}\right)^{\alpha} + \left(\frac{t_{i-1}}{\beta}\right)^{\alpha}\right] \quad (10)$$

Therefore, the likelihood function becomes

$$f(t_1, t_2, \cdots, t_n \mid \alpha, \beta) = \frac{\alpha^n}{\beta^{n\alpha}}\prod_{i=1}^{n} t_i^{\alpha-1} exp\left[-\left(\frac{t_n}{\beta}\right)^{\alpha}\right] \quad (11)$$

Strictly for comparison purposes it is worth noting that the maximum likelihood estimate (MLE) for α is given by

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} ln\left({}^{t_n}\!/_{t_i}\right)} \quad (12)$$

As pointed out by (Bain and Engelhardt, 1991), the MLE for α is biased; an unbiased estimate is given (for the failure-truncated case) by

$$\tilde{\alpha} = \frac{n-2}{n}\hat{\alpha} \quad (13)$$

This bias becomes important for small sample sizes (small n).

3.2 *Analysis in WinBUGS*

The WinBUGS package (Spiegelhalter et al., 1995) was used to perform the MCMC sampling from the joint posterior distribution of α and β and to obtain marginal distributions and summary statistics. WinBUGS was also used to carry out the model validation discussed below.

The likelihood function given by Equ. 11 is not preprogrammed into WinBUGS. The so-called "zeros trick" was used to implement Equ. 11. This trick makes use of the fact that if X ~ Poisson(ϕ), then $e^{-\phi}$ is the probability of seeing X = 0. Thus, one creates a vector of size n, with every component equal to zero. Defining ϕ = -log(likelihood) allows WinBUGS to update the parameters in the likelihood function.

Independent, diffuse priors were used for α and β.[1] This was done strictly to allow the results to be compared with MLEs; a strength of the Bayesian approach is that it allows information about parameters to be encoded into a joint prior distribution for α

[1] WinBUGS requires proper priors. Therefore, an improper prior such as one proportional to α^{-1} cannot be used.

and β. For β, which is a scale parameter determined by the units of time in the problem, a diffuse gamma prior was chosen. For α, which is a shape parameter, one might think that a uniform distribution over a reasonable range (a range of 0.3 to 3 is suggested by (Guida et al., 1989)) would be appropriate. However, it was often the case that the marginal posterior distribution for α was very sensitive to the upper limit in the uniform prior. For this reason, a diffuse gamma prior was used for α, also.

3.3 *Results*

We first examine the failure times given for the "sad," "happy," and "noncommittal" systems given on pp. v-vi of (Ascher and Feingold, 1984). We will present the marginal posterior distributions and summary statistics for α and β, along with summary statistics, and then compare these results to the MLEs. In the following section we discuss Bayesian model validation for this problem.

The figures below show a plot of n(t) versus t for each of these three systems. We expect the points to fall on a straight line if λ is constant, and to be concave and convex for decreasing and increasing λ, respectively.

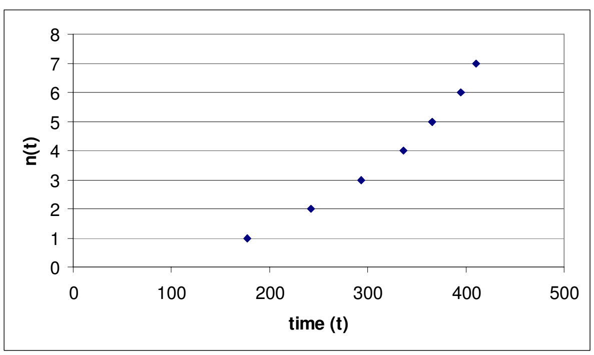

Figure 1 Plot of n(t) vs. t suggests λ increasing with time ("sad" system)

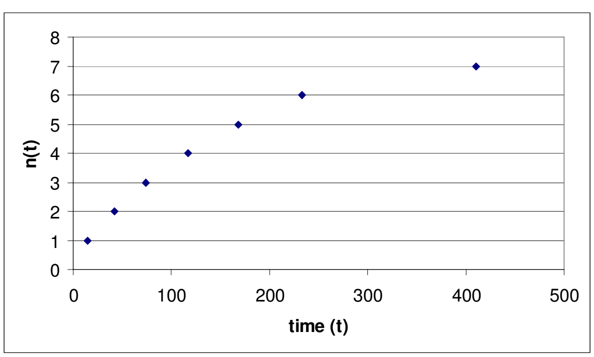

Figure 2 Plot of n(t) vs. t suggests λ decreasing with time ("happy" system)

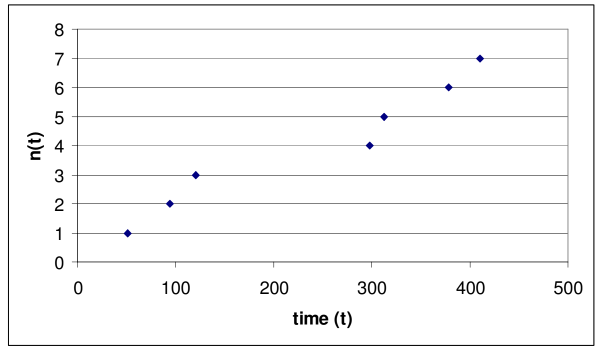

Figure 3 Plot of n(t) vs. t suggests λ constant in time ("noncommittal" system)

For the "sad" system the posterior mean of α is 2.92, with a 90% credible interval of (1.28, 5.12). The marginal posterior distribution of α is shown below. The probability that $\alpha > 1$, implying that λ is increasing with time, is near unity, confirming the "sad" nature implied by Fig. 1 above.

For reference the MLE of α is 3.4, slightly larger than the posterior mean. However, the bias-corrected estimate of α is 2.4, slightly less than the posterior mean. A 90% confidence interval for α is (1.6, 5.1), slightly narrower than the 90% credible interval from the Bayesian analysis.

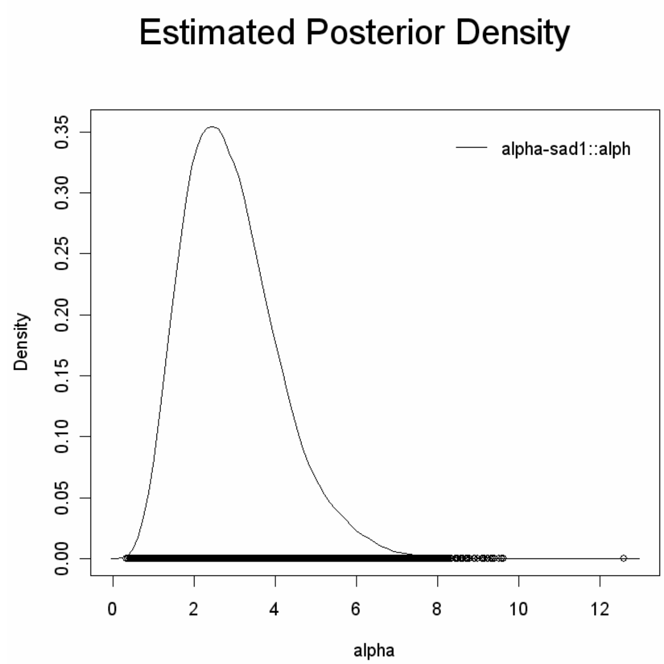

Figure 4 Marginal posterior density of α for "sad" system

For the "happy" system (which is the "sad" system with the interarrival times in reverse order), the posterior mean of α is 0.61, with a 90% credible interval of (0.28, 1.06). The marginal posterior distribution of α is shown below. The probability that $\alpha <$ 1, implying that λ is decreasing with time, is quite large, confirming the "happy" nature implied by Fig. 2 above.

For reference the MLE of α is 0.70, slightly larger than the posterior mean. The bias-corrected estimate of α is 0.50, slightly less than the posterior mean. A 90% confidence interval for α is (0.33, 1.05), slightly narrower than the 90% credible interval from the Bayesian analysis.

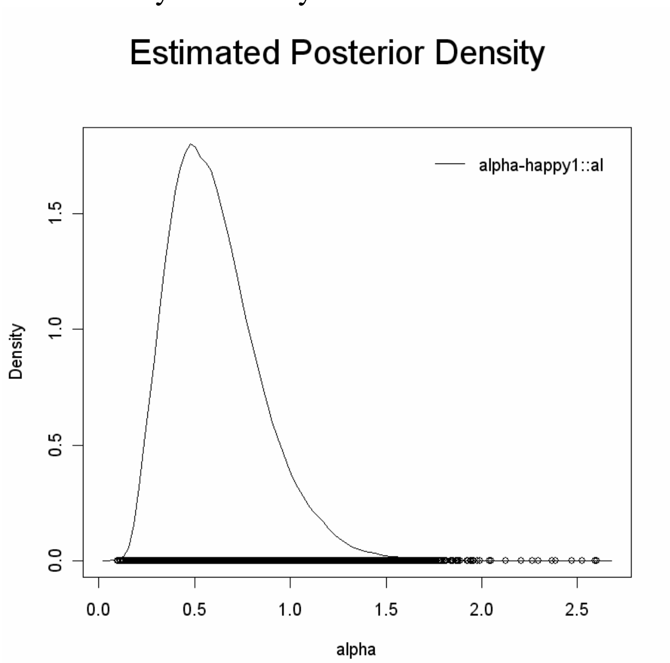

Figure 5 Marginal posterior density of α for "happy" system

Finally, for the "noncommittal" system, the posterior mean of α is 1.09, with a 90% credible interval of (0.48, 1.94). The marginal posterior distribution of α is shown below. The distribution is centered about 1.0, implying that λ is constant with time, confirming the "noncommittal" nature implied by Fig. 3 above.

For reference the MLE of α is 1.28, slightly larger than the posterior mean. The bias-corrected estimate of α is 0.92, slightly less than the posterior mean. A 90% confidence interval for α is (0.60, 1.93), slightly narrower than the 90% credible interval from the Bayesian analysis.

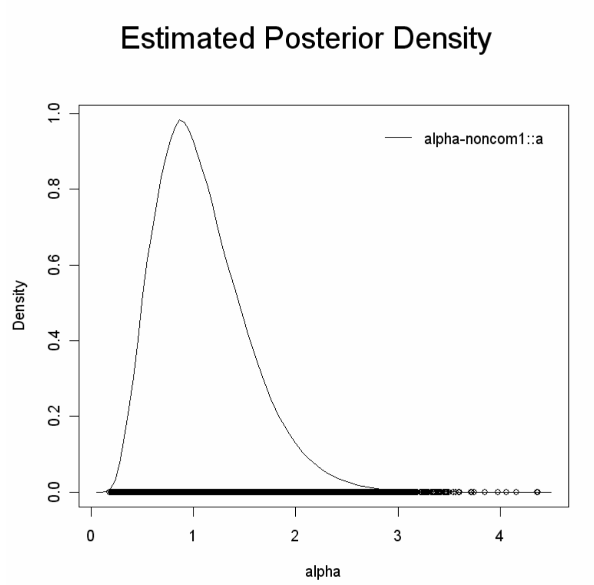

Figure 6 Marginal posterior density of α for "noncommittal" system

4 MODEL VALIDATION

The frequentist approach to model checking or validation typically involves comparing the observed value of a test statistic to percentiles of the sampling distribution for that statistic. Given that the null hypothesis is true, we would not expect to see "extreme" values of the test statistic. The null hypothesis, denoted H_o, is that λ is constant. In terms of the parameters of the power-law process, the null hypothesis is that $\alpha = 1$.

In the hypothesis-testing paradigm, we are required to specify an alternative hypothesis, denoted H_1, and a significance level, <u>before</u> we collect data. If we had no past data on a system's performance, a reasonable alternative would be H_1: $\alpha \neq 1$. This choice allows for the possibility that the system is either getting better or worse with time. We set the significance level at 0.05, the usual value.

As shown in (Bain and Engelhardt, 1991), the quantity $2n\alpha/\hat{\alpha}$ has a chi-square distribution with (2n – 2) degrees of freedom. If the true value of α is significantly less than (greater than) one, we would expect the observed value of $2n/\hat{\alpha}$ to be in the upper (lower) tail of the chi-square distribution. Therefore, we reject H_o: $\alpha = 1$ at a significance level of 0.05 if the observed value of $2n/\hat{\alpha}$ is < $\chi^2_{0.025}(2n-2)$ or > $\chi^2_{0.975}(2n-2)$.

Because of the bias in the MLE for α, it will be harder to detect a "happy" system than a "sad" one.

In our case, we cannot reject H_o: $\alpha = 1$ for the "happy" system at a 0.05 level of significance. The p-value is about 0.13 in this case. For the "sad" system, we can reject H_o at a 0.05 level of significance. As expected, we cannot reject H_o for the "noncommittal" system.

The simplest Bayesian approach to model-checking involves calculating the posterior probability of the various hypotheses and choosing the one that is most likely. However, this is problematic for point hypotheses such as we have here, because $\Pr(\alpha = 1) = 0$, because the posterior distribution is continuous. Qualitatively, one could examine the posterior distribution to see how likely various values of α are. Doing this for the three cases here we would conclude that $\alpha > 1$ for the "sad" system, $\alpha < 1$ for the "happy" system, and $\alpha \approx 1$ for the "noncommittal" system. Quantitatively, $\Pr(\alpha > 1)$ is 0.98 for the "sad" system. For the "happy" system $\Pr(\alpha < 1)$ is 0.93. For the "noncommittal" system, $\Pr(\alpha > 1)$ is 0.53 and $\Pr(\alpha < 1)$ is 0.47.

We can also use summary statistics based on the joint posterior distribution of α and β, as described in (Gelman et al., 2004) to compare a power-law process model with an alternative model, such as an exponential distribution for the failure interarrival times (i.e., constant λ). In this approach, we are quantifying the ability of a model to replicate the observed data. Our choice of summary statistic is motivated by the fact that the expected number of failures in cumulative time t is given by

$$E[X(t)] = \mu(t) = \int_0^t \lambda(s)ds \tag{14}$$

The distribution for the number of failures in time t is Poisson with mean given by Equ. 14. For the exponential distribution, this reduces to simply λt, and for the power-law process we find

$$\mu(t) = \left(\frac{t}{\beta}\right)^{\alpha} \tag{15}$$

We use the observed values of x and t to form the statistic

$$\chi^2_{obs} = \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\mu_i} \tag{16}$$

We then generate replicate values of x from its posterior predictive distribution, and construct an analogous statistic:

$$\chi^2_{rep} = \sum_{i=1}^{n} \frac{(x_{rep,i} - \mu_i)^2}{\mu_i} \tag{17}$$

NB: the observed data are not being used twice, at least not in the sense that empirical Bayes is sometimes criticized for using observed data to both estimate a prior distribution and update that prior for the parameter of interest. Both of these statistics, defined analogously to the frequentist chi-square statistic, have a posterior distribution. χ_{obs}^2 plays the role of the theoretical distribution in the frequentist setting, and χ_{rep}^2 plays that of the summary statistic based on the data; in this case the "data" are replicate values from the Bayesian model. In the frequentist setting, if the summary statistic calculated from the data is in the tail of the theoretical distribution, we are led to reject our model. The p-value is sometimes used to measure the degree to which the data are at conflict with the model. We will adopt that term here, and define the Bayesian p-value to be $\Pr(\chi_{rep}^2 \geq \chi_{obs}^2)$. However, instead of choosing an arbitrary p-value (e.g., 0.05) and rejecting a model with a p-value below the cutoff, we will use the p-value to select the model that is better at replicating the observed data. This will be the model with the larger p-value. The table below shows the Bayesian p-values for the power-law process and exponential model for each of our three systems. Based on these results, we would select the power-law process for the "sad" and "happy" systems, and the exponential model for the "noncommittal" system.

Table 1 Bayesian p-values for alternative models

System	*Power-law process*	*Exponential model*
"sad" ☹	0.60	0.56
"happy" ☺	0.62	0.23
"noncommittal" 😐	0.56	0.71

5 PREDICTING FUTURE FAILURES

In many applications, such as modeling software reliability growth, it is useful to be able to predict number of failures in a future time interval. This can be done easily with WinBUGS, using Equ. 4. WinBUGS generates a number of failures from the posterior predictive distribution, averaging over the joint posterior distribution of α and β. Similarly, it can also be used to predict the time of the $(n + 1)^{st}$ failure.

We illustrate this for "sad," "happy," and "noncommittal" systems we have been analyzing from (Ascher and Feingold, 1984). For all three systems, the last recorded failure was at 410 hours. Figure 7 shows the distribution for the predicted number of failures of the "sad" system during the upcoming 25 hours, that is, between 410 and 435 hours. The most likely number of failures in this 25-hour interval is 0, closely followed by 1 failure. The expected number of failures is 1.35, with a 90% interval of (0, 4).

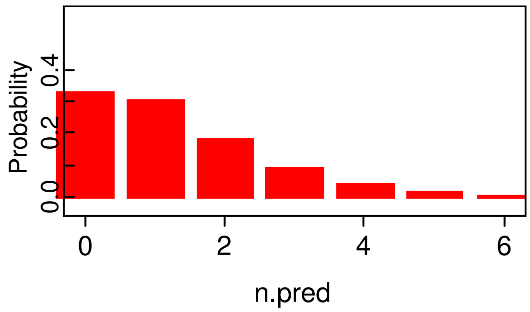

Figure 7 Probability distribution for predicted number of failures between 410 and 435 hours for Ascher "sad" system

Figure 8 shows the distribution of the predicted number of failures between 410 and 435 hours for the "happy" system. The "happy" system is much less likely to fail during this time interval than is the "sad" system. The expected number of failures for the "happy" system is 0.26, with a 90% interval of (0, 1).

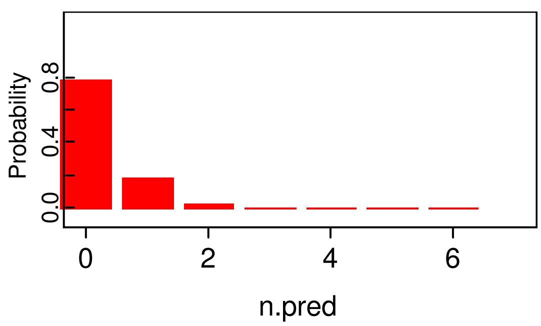

Figure 8 Probability distribution for predicted number of failures between 410 and 435 hours for Ascher "happy" system

For the "noncommittal" system, Figure 9 shows the distribution of the predicted number of failures. As expected, the results are intermediate between the extremes of the "sad" and "happy" systems. The expected number of failures is 0.48, with a 90% interval of (0, 2).

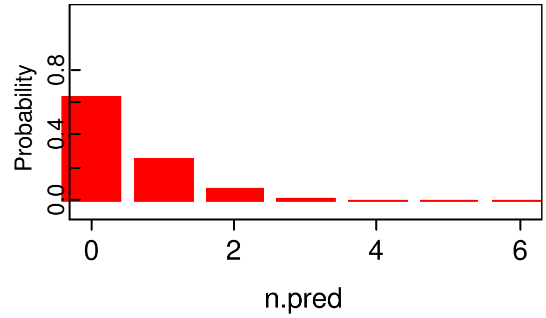

Figure 9 Probability distribution for predicted number of failures between 410 and 435 hours for Ascher "noncommittal" system

6 REFERENCES

1. Harold Ascher and Harry Feingold, Repairable Systems Reliability, Marcel-Dekker, 1984.

2. Lee Bain and Max Engelhardt, Statistical Theory of Reliability and Life-Testing Models, Marcel-Dekker, 1991.

3. Zhao Chen, Bayesian and Empirical Bayes Approaches To Power Law Process and Microarray Analysis, Ph.D. Dissertation, University of South Florida, July 12, 2004.

4. A. Gelman et al., Bayesian Data Analysis, Second Edition, Chapman & Hall/CRC, 2004.

5. M. Guida, R. Calabria, and G. Pulcini, "Bayes Inference for a Non-Homogeneous Poisson Process with Power Intensity Law," *IEEE Transaction On Reliability*, Vol. 38, No. 5, December 1989.

6. D. Spiegelhalter et al., *BUGS: Bayesian Inference Using Gibbs Sampling*, MRC Biostatistics Unit, Cambridge, England, 1994, 2003, www.mrc-bsu.cam.ac.uk/bugs.